

Citation for published version:

Gunkel, DJ & Bryson, J 2014, 'Introduction to the special issue on machine morality: The machine as moral agent and patient', *Philosophy & Technology*, vol. 27, no. 1, pp. 5-8. <https://doi.org/10.1007/s13347-014-0151-1>

DOI:

[10.1007/s13347-014-0151-1](https://doi.org/10.1007/s13347-014-0151-1)

Publication date:

2014

Document Version

Peer reviewed version

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Special Issue - Introduction

David J. Gunkel & Joanna Bryson, Editors

Moral philosophy, although historically defined in different ways by different intellectual traditions, is essentially about intersubjectivity. It concerns how a subject responds to and takes responsibility for their thinking and acting in the face of another—an other who is, in one way or another, recognized as subject to the same moral consideration as the actor. Consequently, moral philosophy concerns—whether it is ever explicitly recognized or not—who (or what) gets to be considered a subject of moral consideration, a moral subject. Moral subjectivity has two recognized aspects: moral agency, the capacity to be responsible for moral actions; and moral patiency, the property of being the recipient of moral obligations. This special issue of *Philosophy and Technology* is one of the first coordinated efforts to consider explicitly both sides of the moral relationship in situations involving artificially intelligent artifacts and machines.

When it comes to deciding these fundamental questions of moral standing, we typically distinguish *who* is a moral subject from *what* is not. And here the two small and seemingly insignificant words of “who” and “what” make, as Jacques Derrida (2005, 80) points out, all the difference. Unfortunately, the distinction between who is a moral subject and what remains a mere object has never been completely resolved. In fact, the history of moral philosophy can be interpreted as a progressive unfolding whereby what had been a mere object, something we do not need to care about in moral terms, comes to be considered another moral subject. “When god-like Odysseus returned from the wars in Troy,” Aldo Leopold (1966, 237) reminds us, “he hanged all on one rope a dozen slave-girls of his household whom he suspected of misbehavior during his absence. This hanging involved no question of propriety. The girls were property. The disposal of property was then, as now, a matter of expediency, not of right and wrong.” During the time of Odysseus (at least as interpreted by Leopold), it was only the male head of the household who had been considered a proper moral subject, everything else, his women, his children, his slaves, were property under his possession,

ostensibly objects that could be used, abused, and disposed of as he saw fit—at least with respect to the legal conventions of that period. Since that time, these previously excluded others—whether slaves, women, foreigners, or children—have been admitted (and often only after considerable struggles and resistance) into the gated-community of moral subjects, becoming other “persons” who count, or are at least deserving of some modicum of protection or respect.

Recently, formal philosophical consideration has been given to extending moral consideration to non-human animals (Singer 1975; Regan 1993) and even the environment (Stone 1974; Birch, 1993), which normally comprise the excluded other of what can be called “the anthropocentric tradition” in ethics. Even in the biological sciences, referring to an animal by name or the use of gendered pronouns (“he” or “she”) or even the generic “who,” is considered a political act, which is excluded, on principle, by many scientific journals (ref.), and more recently included, in some areas of research, at least in the case of non-human primates. Yet while the formal literature often gets bogged down in questions of about the level of consciousness, sentience, or awareness that would be necessary for moral standing, ordinary language and daily behavior indicates that not only the environment but many mundane objects, such as paintings, laws, or cars (Floridi 2008), may deserve status beyond mere possessions of their owners or creators.

Currently we stand on the verge of a potentially revolutionary challenge to many of the standard practices in moral philosophy. We find ourselves faced with another set of entities that have routinely (though not exclusively) been considered nothing more than objects—not *someone* to care about, but *something* to be possessed, used, and disposed of like Odysseus’s slave girls or the animals of multinational agri-business. This other form of otherness is the machine—interactive and learning algorithms that, once programmed and installed, make decisions with little or no human oversight; autonomous robots deployed as weapons on the battlefield, rescuers in disasters, caregivers and assistants in the home or hospital; and the intelligent systems now incorporated into what had previously been mere technological conveniences, like the instruments of transportation, communication, and media. If moral philosophy in the final decades of the twentieth century was defined by the animal question, it is entirely

possible that in the early decades of the twenty-first it will be characterized by the machine question.

Although there have been previous efforts to grapple with this problem, we recognized a need to examine more thoroughly the costs, benefits, reasons and rationales for considering any form of technical entity to be any type of moral subject. In response, we (along with our colleague Steve Torrance) hosted a dedicated symposium of the AISB/IACAP 2012 World Congress, which was convened in conjunction with the centenary celebration of the birth of Alan Turing. The timing and the venue of this event were deliberate. On the one hand, this joint meeting brought together an interdisciplinary group of scholars and researchers: philosophers, behavioral scientists, engineers, computer scientists, social scientists, and legal scholars, etc. This was, in our opinion, crucial insofar as the question of the moral position and status of the machine is not something limited to one discipline, method of investigation, or intellectual tradition. Second, Turing's pioneering work, even if he did not express it exactly the terms we had utilized, was the point of origin for our inquiry and what initially launched our efforts. To put it in rather schematic terms, if one is not, following the stipulations of Turing's "game of imitation" (1951) able to distinguish between another human being and a machine in interactive conversations, what would be our responsibility to this other? In other words, what would happen to our understanding and practice of ethics, if it were possible that the other in social relationships were otherwise than human?

This special issue of *Philosophy and Technology* is made-up of papers from that meeting, competitively selected by the journal's peer-review process. The eight articles presented here cover a wide range of issues that became evident during the course of the symposium. But they are all, in one way or another, an attempt to deal with what we perceive as the two main components of the machine question—moral agency and moral patiency. Questions of moral agency, as Luciano Floridi (2010) points out, involve situations where the machine can be defined as the originator of actions having moral consequences. But deciding whether and to what extent a machine might be considered an agent in this way is an issue that is open to considerable debate, and the articles by Steve Torrance, Johnny Hartz Søraker, David Davenport, and Marc Champaign explore

the various aspects of this issue. Questions of moral patiency, by contrast, involve deciding whether and to what extent machines might be understood as “receivers of moral action” (Floridi and Sanders, 2004) and thereby an entity who possesses what are often called “rights” requiring respect. Again this question is highly contested and the contributions from John Basel, Erica L Neely, Mark Coeckelbergh, and David J. Gunkel pursue its various aspects and complications.

While a single journal issue will not finally resolve the machine question, our goal has been to at least open the door to the question, providing an ontology of moral discourse and charting its range of perspectives and consequences. We therefore advance the modest but nevertheless important objective described by G. E Moore at the beginning of his *Principia Ethica*. “It appears to me,” Moore (2005/1903 xvii) writes, “that in Ethics, as in all other philosophical studies, the difficulties and disagreements, of which its history is full, are mainly due to a very simple cause: namely to the attempt to answer questions, without first discovering precisely what question it is which you desire to answer.” This special issue is offered as one effort to begin to discover the questions that we believe will define the opportunities and challenges of moral thinking in the coming decades.